JOURNAL OF BACTERIOLOGY, Feb. 2002, p. 859–860 Vol. 184, No. 3
0021-9193/02/$04.00+0 DOI: 10.1128/JB.184.3.859–860.2002

Integrative Genetic Element That Reverses the Usual Target Gene Orientation

Sihui Zhao and Kelly P. Williams*

Department of Biology, Indiana University, Bloomington, Indiana 47405

Received 2 August 2001/Accepted 1 November 2001

A genetic element integrating site specifically into a prokaryotic gene usually carries a copy of the 3′ portion of that gene that restores the active gene even as the original is disrupted. A cryptic element in *Mesorhizobium loti* instead carries a copy of the 5′ end of the tRNA gene into which it integrated. This has implications for the evolution of new integrase-site combinations.

Prokaryotic genetic elements often encode integrases that allow site-specific integration into the host chromosome. When the host integration site (*attB*) lies within a gene, the element usually carries at its integration site (*attP*) the same segment of the gene as that displaced; integration disrupts yet smoothly restores the active gene (Fig. 1A). This strategy allows the use of host sequences that are conserved (and therefore reliable) as integration sites without compromising host viability. Although DNA strand crossover occurs in very small (7-bp) segments, elements have apparently been able to capture from a host genome the somewhat larger segment required to restore the target gene. In all previously described cases, it is the 3′ portion of the target gene that is captured in *attP*.

Why are 3′ ends of genes captured in *attP*s rather than 5′ ends? Several explanations have been proposed (2, 10). (i) Regulatory signals at the 5′ ends of genes are more difficult to replace and more crucial than signals at the 3′ ends of genes. An apparent factor-independent transcriptional terminator is frequently found downstream of the gene fragment in *attP*, but a terminator may be simpler to provide, and function in a broader host range, than the promoter that would need to accompany 5′ capture. (ii) A 5′ gene fragment displaced by integration could create an expressed pseudogene that is potentially deleterious either to a virus during lytic growth or to the host after integration. (iii) The mechanism whereby target gene fragments are captured may be directional if, for example, tRNAs or mRNAs mark their own genes by hybridization. Capture may be too infrequent to access experimentally, but enlarging of the database of integration site usage will serve such inquiry.

The rapidly increasing set of completely sequenced bacterial genomes is a rich source of cryptic element sequences. For the majority of classical integrases of the tyrosine recombinase family, a tRNA gene in the host chromosome serves as the target site (7, 10). The integrase gene (*int*) and *attP* are typically adjacent, facilitating their coevolution, and in about half of the cases *int* is situated next to the intact tRNA gene in the integrated element. A simple approach can identify candidate endpoints for some cryptic elements: find an adjacent *int*-tRNA gene pair in an annotated genome and then perform a search with the tDNA sequence for a large duplicated segment in the same orientation as the tRNA gene and across from *int*. In the recently sequenced *Mesorhizobium loti* chromosome (4), we identified two cryptic elements, Mlo45V and Mlo105R (named for the host species, the size in kilobase pairs, and the identity of the tRNA gene), in this way. For these and the previously identified *M. loti* symbiosis island (9), the intact tRNA genes comprising *attL* are oriented toward the element, as the usual 3′-end duplication strategy demands (Fig. 1A).

A fourth integrase gene of *M. loti* is adjacent to a tRNA-Ser gene; however, this tRNA gene aims away from *int*. Nonetheless, a 39-bp block with sequence and orientation identical to that of the 5′ portion of the tRNA gene is found 38 kbp away, on the other side of the integrase gene (Fig. 1B and 2). This preliminary indication for a cryptic element (Mlo38S) requires further support, since its *attP* would be exceptional. Comparison with other partitions of the *M. loti* genome suggests an exogenous origin for Mlo38S (Table 1). As for other *M. loti* genetic elements, the (A+T):(C+G) ratio and codon bias (5)

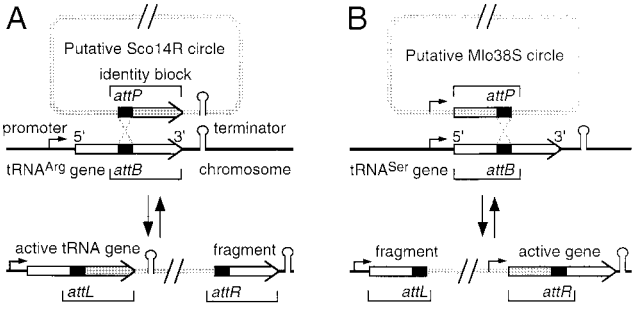

FIG. 1. Restoration of tRNA genes upon integration. (A) Blocks of sequence identity (brackets) between reacting circular DNAs typically extend from the crossover segment (solid boxes) to the 3′ end of the target gene; consequentially, the gene is restored upon integration and aims toward the integrated element. (B) Mlo38S reverses the usual orientation of its target gene by, instead, replacing its 5′ end, which presumably creates an expressed pseudogene. Despite the differing relationships to target genes, the integrases from these two elements are closely related (31% identity). The 14,263-bp element Sco14R was found in *S. coelicolor* genome sequence data (www.sanger.ac.uk); annotation of its gene content (GenBank entries AL035707 and AL049573) suggests a close relationship to integrative plasmid pSAM2.

* Corresponding author. Mailing address: 1001 East Third St., Bloomington IN 47405. Phone: (812) 856-5697. Fax: (812) 855-6705. E-mail: kwilliam@bio.indiana.edu.

Mlo38S *attL* (*M. loti* bp 310039-310179) acaggcctttgccgggcacactagagcagtcgaatgtgatggacagtcact
GGAGGGATGGCCGAGCGGTTTAAGGCACCGGTCTTGAAAttcgctcgctgatgtacctcccctccccgttccatgctgccaagtccttgt

Mlo38S *attR* (*M. loti* bp 347591-347827) *int*…CGCGGCAGAGCAAGTTGACGTGGCATTGCAGGCGGCTATAAGCAGCG
AGCGAAAAGGCAAATAGttgcaaagtgcgtagcaaagtggtttttcggccccggcgccggcaaacaaaagcgcaatgattgcaactagtt
GGAGGGATGGCCGAGCGGTTTAAGGCACCGGTCTTGAAAACCGGCGTGGGCGCAAGTTCACCGTGGGTTCGAATCCCACTCCCTCCGCCA

B. japonicum AF200320 *int*…TAGcaaagcggtagcagagtccacggaacgacgaaaagctaaacaatttcaaggctt
GGAAGGGTGGCCGAGTGGTTTAAGGCACCGGTCTTGAAAACCGGCGTGCCTGCAAGGGTACCGTGGGTTCGAATCCCACCCCTTCCGCCA

FIG. 2. Mlo38S ends and a *Bradyrhizobium* homolog. tDNA-Ser and integrase gene sequences are in uppercase. Anticodon loop tDNA is underlined. Wavy underlining indicates plausible −35 and −10 promoter elements. Sequence data for the presumed *attL* in *B. japonicum* is not available.

of Mlo38S are distinctly elevated relative to those of the remainder of the genome. Mlo38S has the lowest percentage of genes with at least some functional assignment; other than its integrase, only two of its proteins are assigned (4), homologs of the two subunits of the phage P22 enzyme that initiates headful DNA packaging. This suggests that Mlo38S may be a prophage, although the source of virion structural proteins has not been identified.

A hypothetical problem arising from integration that replaces the 5′ end of an expressed gene is that the gene fragment remaining at *attL* would be expressed (2). The transcriptional status of Mlo38S *attL* is unclear: potential −35 and −10 sequences indicated in Fig. 2 have unusually long spacing; a better candidate for a strong promoter would initiate transcription 426 bp upstream of the tDNA fragment. It is also unclear whether expression of such a tDNA fragment would, in fact, be deleterious, but if its expression has been dampened by mutation since the integration of Mlo38S, then the host may not survive with the weakened gene that would be left after excision of the element. The tRNA-Ser encoded at Mlo38S *attR* must be expressed, being the only *M. loti* tRNA that can decode UCA codons; its likely promoter is at the end of the integrase gene (Fig. 2).

Bradyrhizobium japonicum provides a redundant instance of the Mlo38S integration system; it has the gene for the closest known BLAST homolog of the Mlo38S integrase (52% identity), at the same orientation to the same tRNA gene (Fig. 2). Phylogenetic analysis (10; see also the Tyrosine Recombinase Web Site at http://members.home.net/domespo/trhome.html) tends to place these two integrases, associated with a 5′ gene fragment capture event, into a small subfamily with integrases associated with the standard capture of the 3′ ends of different tRNA genes, from pSE101, pSE211, and the element Sco14R (Fig. 1A) that we have newly identified in *Streptomyces coelicolor*.

TABLE 1. Partitions of the *M. loti* genome

Partition	Size (bp)	A+T:C+G ratio	No. of genes	% Assigned	Codon bias	*attLR* shared tDNA (bp)
Plasmid pMLa	351,911	0.687	320	50.0	0.408	
Plasmid pMLb	208,315	0.670	209	35.9	0.383	
Symbiosis island	610,992	0.673	580	62.4	0.398	17 (3′)
Mlo45V	45,239	0.638	52	19.2	0.294	48 (3′)
Mlo105R	105,450	0.690	121	15.7	0.368	48 (3′)
Mlo38S	37,638	0.659	43	7.0	0.325	39 (5′)
Remainder	6,236,755	0.584	5,956	55.1	0.197[a]	

[a] Evaluated after omitting the 90 reference reading frames (5), all contained in this partition; the value is 0.194 without the omission.

While some integrases appear to promote crossover at T-loop or T-stem/acceptor-stem junction tDNA (1, 10), the best-studied tDNA-specific integrases catalyze DNA strand crossover in a 7-bp segment corresponding precisely to anticodon-loop tDNA (3, 6, 8). In these cases, as for pSE101, pSE211, and Sco14R, the blocks of identical sequence shared by *attL* and *attR* just encompass the anticodon-loop tDNA and extend 3′. Crossover at anticodon-loop tDNA, which is flanked by the symmetrical sequences that encode the anticodon stem, mirrors the preference of lambda integrase (whose *attB* is not in tDNA) for symmetrical core binding sites flanking its crossover segment. The *attLR* identity block for Mlo38S also just encompasses anticodon-loop tDNA, although extending 5′ (Fig. 2). Termination of identity at the commonly used anticodon-loop sublocation, and the close relationship of its integrase to others associated with 3′ captures, suggests that Mlo38S exhibits a rare outcome of a standard mechanism of target gene fragment capture. This mechanism can now be said not to oblige the 3′ direction. The strong directional bias that is still observed may yet be imparted upon capture or may result from negative selection after 5′ captures.

REFERENCES

1. **Auvray, F., M. Coddeville, R. C. Ordonez, and P. Ritzenthaler.** 1999. Unusual structure of the *attB* site of the site-specific recombination system of *Lactobacillus delbrueckii* bacteriophage mv4. J. Bacteriol. **181:**7385–7389.
2. **Campbell, A. M.** 1992. Chromosomal insertion sites for phages and plasmids. J. Bacteriol. **174:**7495–7499.
3. **Hauser, M. A., and J. J. Scocca.** 1992. Site-specific integration of the Haemophilus influenzae bacteriophage HP1. Identification of the points of recombinational strand exchange and the limits of the host attachment site. J. Biol. Chem. **267:**6859–6864.
4. **Kaneko, T., Y. Nakamura, S. Sato, E. Asamizu, T. Kato, S. Sasamoto, A. Watanabe, K. Idesawa, A. Ishikawa, K. Kawashima, T. Kimura, Y. Kishida, C. Kiyokawa, M. Kohara, M. Matsumoto, A. Matsuno, Y. Mochizuki, S. Nakayama, N. Nakazaki, S. Shimpo, M. Sugimoto, C. Takeuchi, M. Yamada, and S. Tabata.** 2000. Complete genome structure of the nitrogen-fixing symbiotic bacterium Mesorhizobium loti. DNA Res. **7:**331–338.
5. **Karlin, S., J. Mrazek, and A. M. Campbell.** 1998. Codon usages in different gene classes of the Escherichia coli genome. Mol. Microbiol. **29:**1341–1355.
6. **Pena, C. E., J. M. Kahlenberg, and G. F. Hatfull.** 2000. Assembly and activation of site-specific recombination complexes. Proc. Natl. Acad. Sci. USA **97:**7760–7765.
7. **Reiter, W. D., P. Palm, and S. Yeats.** 1989. Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. Nucleic Acids Res. **17:**1907–1914.
8. **Smith-Mungo, L., I. T. Chan, and A. Landy.** 1994. Structure of the P22 att site. Conservation and divergence in the lambda motif of recombinogenic complexes. J. Biol. Chem. **269:**20798–20805.
9. **Sullivan, J. T., and C. W. Ronson.** 1998. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. Proc. Natl. Acad. Sci. USA **95:**5145–5149.
10. **Williams, K. P.** Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. Nucleic Acids Res., in press.